DOCUMENT RESUME

ED 314 496 TM 014 419

AUTHOR Prosser, Barbara
TITLE Beware the Dangers of Discarding Variance.
PUB DATE Jan 90
NOTE 24p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 25-27, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; Data Analysis; *Multiple Regression Analysis; Predictor Variables; *Research Design; Research Methodology; Research Problems
IDENTIFIERS Planned Comparisons; *Variance (Statistical)

ABSTRACT

The value of variance is emphasized, and the element of design, frequently not adequately understood, is clarified to underscore the importance of variance to the researcher. Two analytic methods, analysis of variance (ANOVA) and multiple regression, are discussed in terms of how each uses/applies variance. Advantages and major difficulties with ANOVA are described, with the main problem with ANOVA and related methods being that they require predictor variables to be nominally scaled. The strengths of multiple regression are presented, and the use of planned comparisons is discussed. Discarding variance is cited as one of the prevalent methodology errors in education dissertations. The tendency to equate ANOVAs with experimental designs leads to the discarding of variance through the process of putting continuous variables into categories. A small data set of six groups illustrates how keeping all variance can help provide a more effective and sensitive analysis of data. One data table and a 26-item list of references are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED314496

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BARBARA PROSSER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEWARE THE DANGERS OF DISCARDING VARIANCE

Barbara Prosser

University of New Orleans

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 25, 1990.

TM014119

BEST COPY AVAILABLE

ABSTRACT

The value of variance is emphasized in the present study. The element of design, frequently not well enough understood, is clarified to underscore the importance of variance to the researcher. Two analytic methods, ANOVA and multiple regression, are discussed in terms of the use each makes of variance. Major difficulties with ANOVA are described, and strengths of multiple regression are given in contrast. Discarding variance is cited as one of the prevalent methodology errors in education dissertations. A small data set is utilized to show how keeping all variance can be used to good result: a more effective and sensitive analysis of data.

Variance is at the core of all quantitative research. It is the "stuff" of which such studies are made (Thompson, 1988a). Appearing in a number of forms in research and technical studies, variance must be well understood if one is to comprehend the literature and to design and do sound research. Kerlinger (1986) advises that this important concept be mastered thoroughly in terms of not only its characteristics and purposes, but also its notions and manipulations. Variance is called a concept "of the highest importance in research and in the analysis of research data" (Kerlinger, 1986, p. 84). Pedhazur (1982, p. 5) asserts "that the scientist is essentially interested in explaining variance. Variability in the behavioral sciences is in itself the object of much curiosity and interest."

Despite the recognized importance of variance, some researchers in the behavioral sciences choose techniques which discard it by categorizing variables. This paper cautions against such practice and discusses the advantages of using analytic techniques which do not throw away information that has been gathered with painstaking care and planning. Doctoral students, in particular, are warned to avoid the pitfall of discarding variance in their dissertations as well as their future research efforts (Goodwin & Goodwin, 1985).

Distinguishing Between Design and Analysis

Good quality in research is the result of the researcher's commitment to a series of careful choices during each and every

phase of the investigation, from design to analysis. Selectivity at the beginning tends to beget selectivity at the end. As Winer (1962, p. 2) notes, design is everything to a study:

> Two experiments having identical objectives may be designed in quite different ways; at the same cost in terms of experimental effort, one design may lead to unambiguous results no matter what the outcome, whereas the second design could potentially lead to ambiguous results no matter what the outcome.

Concerns regarding the importance of designs have continued to have their lengthy history, just as have those focusing on analysis of data (Thompson, 1985). Although design and analysis were historically viewed as separate and distinct issues, the line between them began to blur. Unfortunately, design is now all too often equated with analysis. Thompson (1981) has investigated the origin of this confusion and has suggested that the prevalence of ANOVA has been a major contributor to the problem. ANOVA methods became, in time, identified with the power of experimental design, and vice versa (Thompson, (1985, p. 50).

This tendency to equate ANOVA with experimental designs leads to the discarding of variance through the process of putting continuous variables into categories. Thompson (1981, p. 5) observes:

> Researchers tend to prefer experimental over correlational research designs because experimental designs provide more complete information about causality. Why does this situation contribute to OVAism? Because some researchers confuse research design with the statistical techniques which are used to analyze the data which the designs help to generate.

Thus, having a clear understanding of the importance of design, and its separateness from analysis, enables the researcher to make more thoughtful decisions in planning studies. It also affords an accurate evaluation of variance in data analysis.

ANOVA

"OVA," as used in this paper, is a collective term for analysis of variance (ANOVA) and its analogs: ANCOVA, MANOVA, and MANCOVA. OVA methods are prevalent in social science research--the "tools of choice" for many (Thompson, 1988, p. 3). First introduced in 1925 by Sir Ronald Fisher in a work titled Statistical Methods for Research Workers, ANOVA was intended for use with experimental designs. However, it began to be used extensively by educational researchers to analyze data with both experimental and nonexperimental predictors (Lopez, 1989). The popularity of OVA methods grew; in the peak decade prior to

1980, OVA methods comprised approximately 41 per cent of the statistical techniques used in articles published in the American Educational Research Journal (Willson, 1980).

Contributing to the popularity and frequent use of the OVA techniques were three perceived advantages. In the 1920's, ANOVA offered quicker computation--a feature more important in the age prior to computers. Now, computers are widespread and analyze data with assured speed and accuracy; difference in computation time is just a matter of milliseconds (Knapp, 1978; Thompson, 1985). A second advantage noted was neatness in interpretation, thanks to perfectly uncorrelated omnibus tests that arise in balanced designs--i. e., there are equal numbers in each cell of a study. Results are indeed cleanly interpretable, but at the considerable price of discarding data in order to establish and then maintain balance throughout the study. Finally, there was the advantage of testing interaction effects. OVA methods do permit, through the partitioning of variance, the researcher to interpret the interaction effects of two or more independent variables as they affect the dependent variable. However, interaction effects can also be studied using non-OVA techniques that do not require specialized conditions like balance and partitioning. Kerlinger and Pedhazur (1973), as well as Thompson (1985), have demonstrated that interaction effects can be studied through the use of product variables in multiple regression analysis. Therefore, in nonexperimental research, OVA methods

have advantages that are of dubious value. The researcher must be clearly aware of these limitations.

There are, in addition, seriously disturbing aspects about the use of OVA methods in nonexperimental research. Their misuse can compromise the integrity of an entire study. This is true mainly because of the specialized conditions required by the OVA's.

The main problem with OVA methods is that they require predictor variables to be nominally scaled. In many cases, this necessitates the "squandering [of] much information" (Cohen, 1968, p. 441). The trouble with this is that most variables, other than experimental conditions, are higher than nominally scaled. Thompson (1986) notes that one exception to this is biological sex of the subjects in a study. Typically, though, the use of sex as an independent variable is not justified theoretically; sex information is merely convenient to collect and then analyze. It often just facilitates measurement. In criticizing these specialized requirements of the OVA methods, Thompson (1981, p. 8) argues that

> When we reduce interval level of scale data to the nominal level of scale, we are doing nothing less than thoughtlessly throwing away information which we previously went to some trouble to collect. If research is conducted for the purpose of acquiring knowledge, then is it consistent with our purpose

> to employ a method which "throws away" information which might provide a more refined understanding of the phenomena which we are studying?

This feature of forcing intervally scaled predictor variables to the nominal level has been criticized by others. Pedhazur (1982) notes that the practice of making continuous variables fit into categories is used too often in the social sciences. He attributes some of the conflicts present in the research literature to this practice. "Categorization leads to a loss of information, and consequently to a less sensitive analysis" (Pedhazur, 1982, pp. 452-453).

Cliff (1987, p. 130) also argues against discarding variance on intervally scaled predictor variables in order to perform OVA analyses because it creates problems in almost every case:

> Such divisions are not infallible; think of the persons near the borders. Some who would be highs are actually classified as lows, and vice versa. In addition, the "barely highs" are classified the same as the "very highs," even though they are different.

Siebold and McPhee (1979, p. 355) point out that converting variables to nominal scale distorts relationships among the independent variables:

> Advancement of theory and the useful application of

> research findings depend not only on establishing that a relationship exists among predictors and the criterion, but also upon determining the extent to which those independent variables, singly and in all possible combinations, share variance with the dependent variable. Only then can we fully know the relative importance of independent variables with regard to the dependent variable in question.

Humphreys and Fleishman (1974, p. 468) warn the investigator using OVA methods in a nonexperimental design and categorizing individual differences like scores that pretending the variables are defined under experimental control will not make the pretense true. Such analysis "not infrequently produces in both the investigator and his audience the illusion that he has experimental control over the independent variable. Nothing could be more wrong." With this problem of a false sense of control comes a tendency on the part of the investigator to interpret statistically significant relationships causally.

Kerlinger (1986 p. 558), too, advises strongly against the practice of forcing intervally scaled variables to nominal scale, explaining that it can cause results to be misinterpreted:

> Partitioning a continuous variable into a dichotomy or trichotomy throws information away....To reduce a set of values with a relatively wide range to a dichotomy is to reduce its variance and thus its

possible correlation with other variables. A good rule of research data analysis, therefore, is: Do not reduce continuous variables to partitioned variables (dichotomies, trichotomies, etc.) unless compelled to do so by circumstances or the nature of the data (seriously skewed, bimodal, etc.).

A second concern with OVA methods is that even after potentially valuable information is discarded when continuous variables are put into categories, more data still may be lost as the researcher seeks to maintain a balanced design. Maintaining balance, or equal numbers in each cell, is another special requirement of the OVA methods. This assures that effects will be orthogonal. No correlation among the predictor variables in explaining the dependent variable will be present, and results will be conveniently boxed for a "clean" interpretation.

Several problems stem from this requirement of balance. First, preserving a balanced design throughout an entire research is often very difficult. In reality, subjects seldom come in presized packages. Human subjects, for instance, often drop out of studies--throwing cells out of proportion and signaling the demise of still more data as the researcher makes adjustments to achieve balance once again (Thompson, 1986). Managerial and perhaps even ethical dilemmas may present themselves, and the

assumption that OVA methods offer simplicity is recognized as an illusion. Forcing balance to keep predictor effects orthogonal is misleading in the social sciences, where correlation among variables is common.

Still another problem resulting from the requirement of balanced, orthogonal variables that are nominally scaled is decreased reliability. This tends to reduce power against Type II error (Cohen, 1968; Thompson, 1986). As Cliff (1987, p. 130) notes, "Reducing a reliable variable to a dichotomy makes the variable more unreliable, not less." Theoretically, statistically significant results are possible only when variables are reliably measured. Reliability, however, is a function of variance. Thompson points out that longer tests usually have more reliability than shorter tests because they have more test items allowing subjects to get scores that are more widely distributed. In OVA methods, then, the discarding of variance--information about those predictor variables that started out higher than nominally scaled--leads to lessened reliability and increased Type II error probability.

The irony of the situation is that many are under the erroneous impression that OVA methods offer more power and, therefore, more protection against the Type II error. In reality, power is lessened; research workers run a higher risk of failing to reject a false null. This is indeed a serious error for any researcher to commit; it leads up a blind alley and may

cause a waste of time and effort by generating additional research leading nowhere. Lines of inquiry that otherwise would be important could be abandoned.

OVA methods are called "awkward" and "inefficient" by Humphreys and Fleishman (1974, p. 464). Fortunately, their use in nonexperimental research appears to be decreasing. Thompson (1986) traces this trend, noting that in prominent educational research journals surveyed in 1985, 37 per cent of the articles published between 1979-1983 used OVA techniques--a drop from the previously noted 41 per cent from 1969-78. A further lessening in popularity is seen in the 25 per cent observed in three educational research journals during 1978-87 (Daniel, 1989; Elmore & Woehlke, 1988). However, despite their widely recognized disadvantages, OVA methods continue to be used with marked frequency.

Multiple Regression

Classical multiple regression analyzes various measures by means of traditional correlation analysis. Unlike OVA methods, multiple regression does not require converting independent variables from interval to nominal scale. Substantively important information is not thrown away. Understandably, then, it is not surprising that in some research situations, multiple regression and other general linear methods have been shown to be superior to OVA methods in accurately estimating explained dependent variance (Daniel, 1989; Thompson, 1986). Kerlinger and Pedhazur

(1972, p. 3) commented a long time ago on the attributes of multiple regression:

> It can be used equally well in experimental or non-experimental research. It can handle continuous and categorical variables. It can handle two, three, four, or more independent variables...Finally, as we will abundantly show, multiple regression analysis can do anything the analysis of variance does--sums of squares, mean squares, F rations--and more.

When further compared with the OVA methods, multiple regression has several additional advantages. The researcher does not have to be concerned with a balanced design. This affords both flexibility and reliability that is lacking in analysis of variance L (Darlington, 1968; Thompson, 1988). Furthermore, the use of non-OVA methods does not inflate the possibility of a Type II error. Finally, multiple regression does not distort distribution shapes of or relationships among independent variables (Cohen, 1968; Thompson, 1981).

Multiple regression should be considered by researchers looking at non-OVA methods for analyzing data. The analysis of variance has been used more frequently and has been more popular than it has deserved to be. When the two methods are compared in various applications, multiple regression is strongly favored (Humphrey & Fleishman, 1974). It is, for example, preferred in

most research involving individual-difference variables. While OVA results are easier to interpret, multiple regression gives more accurate analysis as well as increased probability of obtaining statistical significance.

Multiple regression analysis can evaluate the individual and combined effects of predictor variables. The researcher may consult (a) beta weights, (b) bivariate correlations between predictors and the dependent variable, and (c) structure coefficients (Thompson & Borrello, 1985). Contrary to the belief of many researchers, multiple regression can be used to study both linear and curvilinear and main and interaction effects (Thompson, 1989). Its versatility, which is receiving increased attention and appreciation, helps to make it very useful in behavioral science applications.

Dissertation Methodology

Thompson (1988) studied a sampling of doctoral dissertations to identify common methodology errors in them, with the aim of helping others to avoid the same pitfalls. The seven mistakes he found are described in detail, with actual, concrete examples cited from 10 dissertations finished since 1985. The information provided is useful not only to doctoral students and their faculty, but also to all authors of published research.

One of the seven errors described concerns the discarding of variance. Thompson (1988, pp.14-19) warns that

> Dissertations should reflect the recognition that

> discarding variance to conduct chi-square or OVA analyses can lead to serious distortions in interpretations, and that even when OVA methods are appropriate the methods should usually be implemented using regression approaches.

Dissertation students, according to Thompson (1988), frequently discard variance to conduct Pearson chi-square contingency table tests or OVA methods--the OVA methods being more often applied. Distortions are thus introduced with variables that were at first intervally scaled.

Furthermore, to isolate which means are significantly different within OVA ways having more than two levels, a lot of researchers use unplanned (post hoc) multiple comparison tests (e. g., Sheffe, Tukey, or Duncan). Unplanned comparison tests are developed after the results of the study are obtained--not by any prior rationale. This leads post hoc comparisons to be referred to by textbook authors with terms such as "fishing expeditions," "data snooping," and "milking" (Thompson, 1988).

Planned comparisons, also called focused or a priori tests, provide a valuable alternative to post hoc analyses. Typically, planned comparisons involve weighting data by "contrast" sets such as those presented in Table 1. Other kinds of contrasts test for trends in means (Thompson, 1988).

INSERT TABLE 1 ABOUT HERE

Typically, contrasts are developed to sum to zero, as do all five of the contrasts for the data in Table 1. The data represent a hypothetical validity study to determine if various clinical groups score differently on a psychological measure. Like the hypotheses they test or represent, the contrasts are uncorrelated. They sum to zero, as also do the cross-products of each pair of contrasts. Thus, the contrasts given in Table 1 are orthogonal (Thompson, 1988).

Not all researchers believe that planned comparisons need to be orthogonal. They feel that it makes little, if any, difference if f tests were uncorrelated (Winer, 1971). Proponents of orthogonal comparisons believe that there is value in the independence of their inferences, which eliminates some potential difficulty in interpreting the different outcomes.

Planned comparisons are usually preferred to unplanned ones for two reasons: (a) more power against the Type II error--thus more sensitivity to statistical significance; and (b) the tendency "to force the researcher to be more thoughtful in conducting research" (Thompson, 1988, p.18). Planned comparisons are formulated before the data collection, and only a limited number of them can be used in a given study. They can be a motivating force, allowing more worthwhile focus in contrast to the diffusion seen with post hoc tests. While post hoc comparisons certainly have some value, planned comparisons have more scientific importance (Thompson, 1988).

Doctoral students employing OVA methods in dissertation research should be aware of the cautions surrounding their use. Dissertations are evidence of the best work and interests of not only students, but also doctoral faculty. The importance of having dissertations of good quality cannot be underestimated by a university. In most states, boards of higher education employ program review teams to study dissertations thoroughly. Findings of these teams reflect whether or not the university has a sound training program which should be continued, and whether or not its doctoral program graduates will likely be productive scholars (Thompson, 1988).

Discussion

Keeping all variance is vital to a sensitive, effective analysis of data. This principle must be emphasized to all educational researchers, who should make informed choices in regard to every phase of their work, including which methods to use for the analysis of their data. Practices and decision making affecting many will be influenced by research findings (Thompson, 1986).

OVA methods, which first offered the advantage of ease of computation, have been employed too often in nonexperimental research--a use for which they were not intended. Although their popularity in educational research journals appears to be declining, OVA methods continue to be prevalent; they come with a high price in terms of research quality. Their popularity has

contributed to the confusion many now have about design and analysis. OVA methods have been erroneously identified with the power of experimental design. As a result, there is a tendency to discard variance through the process of putting continuous variables into categories, and then further requiring balanced cell sizes. Potentially valuable data are thrown away, and remaining information is put into arbitrary shapes that distort relationships. Findings have decreased reliability, as well as less protection against the Type II error.

OVA methods are appropriate to use only under certain conditions. First, all predictors are inherently nominal. Second, all ways have only two levels. However, such a case would be unusual, though not impossible.

Multiple regression is superior to the OVA methods in research in the social sciences. Without the specialized conditions of the OVA methods, multiple regression retains all variance and results in more reliable and sensitive findings. Its advantages and versatility are becoming more widely recognized by behavioral science researchers.

References

Carver, R. (1978). The case against significance testing. Harvard Educational Review, 48, 378-98.

Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Daniel, L. (1989, January). Commonality analysis with various parametric measures. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.

Edgington, E. S. (1974). A new tabulation of statistical procedures in APA journals. American Psychologist, 29, 25-26.

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Goodwin, L. D. & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Humphreys, L. G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-

differences variances. Journal of Educational Psychology, 66, 464-472.

Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. Journal of Educational Psychology, 66, 873-876.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart, & Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance-testing system. Psychological Bulletin, 85, 410-416.

Lopez, K. (1989, November). Testing interaction effects without discarding variance. Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Siebold, D. R. & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analysis. Human Communication Research, 5, 355-65.

Thompson, B. (1981, November). The problem of OVAism. Paper presented at the annual meeting of the Mid-South Educational

Research Association, Lexington, KY. (Order document #03980, National Auxillary Publication Service, P. O. Box 3513, Grand Central Station, New York, NY 10017)

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B. (1986). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928.

Thompson, B. (1988, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)

Thompson, B. (1988a). Discarding variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21(1), 3-4.

Thompson, B. (1988b). Misuse of chi-square contingency table test statistics. Educational and Psychological Research, 8, 39-49.

Thompson, B. (1989, January). Heuristics for understanding the concepts of interaction, polynomial trend, and the general linear model. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 304 453)

Thompson, B. & Borrello, G. M. (1985). The importance of

structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9(6), 5-10.

Winer, B. J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.

Table 1

Hypothetical Validity Study Data

Group	ID	DV	C1	C2	C3	C4	C5
1	1	10	-1	-1	-1	-1	-1
	2	20	-1	-1	-1	-1	-1
2	3	10	1	-1	-1	-1	-1
	4	20	1	-1	-1	-1	-1
3	5	10	0	2	-1	-1	-1
	6	20	0	2	-1	-1	-1
4	7	10	0	0	3	-1	-1
	8	20	0	0	3	-1	-1
5	9	10	0	0	0	4	-1
	10	20	0	0	0	4	-1
6	11	25	0	0	0	0	5
	12	35	0	0	0	0	5

(Thompson, 1988, p. 46)